Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

STANDARD TESTS IN ENGLISH

S. A. COURTIS
Liggett School, Detroit, Michigan

Standard tests in education are rulers for the measurement of educational products; tools for research work. They should be differentiated from the ordinary examinations on the one hand and the more rigid and formal measurements of the experimental psychologists on the other. An examination, for instance, is supposed to show whether or not the individual has successfully completed a prescribed course of work. Examinations are complex, are usually made afresh each time they are given, and only in the most general sense may the conditions under which they are given said to be controlled.

The tests of the psychologists are "pure science" tests. The test itself often deals with a single elemental phase of mental action, and the conditions under which the test is given are rigidly controlled. For this reason the learning of nonsense syllables, crossing out of *A*'s, and other tests of this character, while of value from the point of view of scientific study of the mind, have little connection with actual classroom work.

Day by day, however, the school is attempting to make certain well-defined changes in the minds and habits of the children in its care. The character and the amount of these changes is not determined by chance, but in accordance with laws which, in their operation, are as constant as those which determine other phenomena of the natural world. Until these laws are known, scientific control of the efficiency of the educational process will be impossible. The greatest need of the educational world today, from the point of improvement of the efficiency of teaching, is scientific yet *practical* measures of the fundamental abilities actually being developed by school work. For it is the possession of such measures that makes possible the study of the development of these abilities from grade to grade and the evaluation of the effects of varying conditions in method.

In the writer's thinking, the functions of standard educational tests are four: (*a*) to secure information that will enable school authorities to formulate in objective terms the ends to be attained in any educational process; (*b*) to measure the efficiency of methods designed to produce the desired results; (*c*) to determine the factors and laws which condition learning and teaching; (*d*) to furnish data that will enable comparisons of school with school, and teacher with teacher for purposes of supervisory control to be made upon *scientific*, *impersonal*, *objective* bases.

If the reader will keep these purposes in mind during the discussions of the tests described below, he will not be led into making unwarranted criticisms. For the tests are neither examinations, classroom exercises, nor measures of elementary psychological activities.

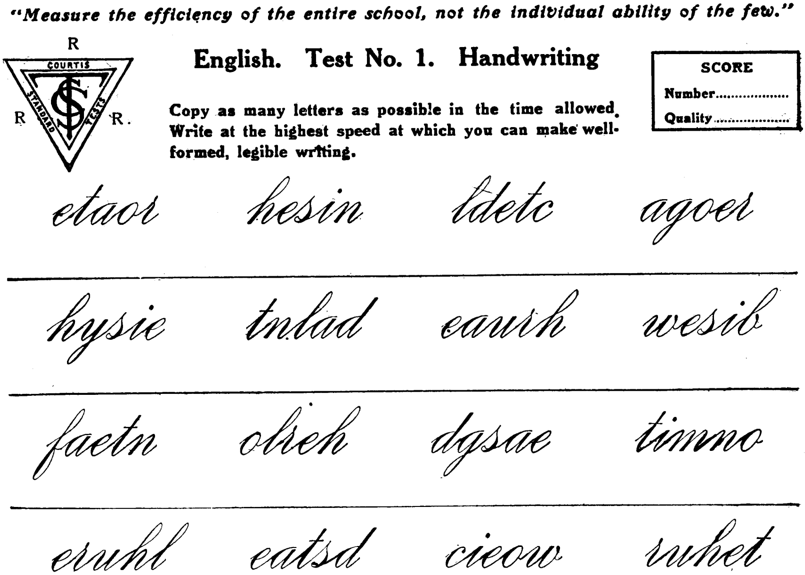
As the readers of this journal know, the writer has for some years been measuring the efficiency of teaching in arithmetic. The development of the abilities involved in work with whole numbers in the four processes has at last been brought under control; but little success has been achieved in the control of the reasoning processes. Recent experimental studies have seemed to show that the chief difficulties are closely connected with the abilities involved in reading. It has become imperative, therefore, to determine standard rates of reading and of comprehension; but in giving tests for this purpose it was found that the material secured could be scored in so many other ways of value that the group of tests has been combined into a single series under the head of "English Tests." These have been issued on the same co-operative basis followed in the arithmetic work, and are now being standardized by teachers and superintendents in different parts of the country. Definite standards cannot be formulated for some time, but it was thought that a summary of the results obtained from a single school would furnish a concrete illustration of the products of such tests and lead to a more widespread interest and a greater degree of co-operation in the work of standardization.

The tests are six in number and have been named as follows: Test 1, "Handwriting"; Test 2, "Dictation"; Test 3, "Original Story"; Test 4, "Normal Reading"; Test 5, "Careful Reading";

Test 6, "Reproduction." They are not distinct tests, but a series of related exercises grouped about a central theme.

The test sheets are printed on paper of a uniform size and bound with appropriate record sheets in pamphlet form. Exactly the same tests are given to all grades from the fourth through high school and university, and that conditions may be kept uniform, instructions for giving, scoring, and tabulating the tests have been printed in convenient form together with necessary record sheets, answer cards, and graph sheets. The tests themselves contain many devices tending to make scoring easy and to reduce the time required to a minimum. They will be supplied at cost to those who will return their results for tabulation with those of other schools. Any teacher of English, even though without previous training in educational measurement, may, at nominal cost, take part in the co-operative investigation with benefit to himself.

The series of tests is based upon the picture of a child, startled by the appearance of a hungry dog while sitting in a doorway drinking from a bowl of milk. The dictation exercises (Test 2) give the first part of a story (Part A) based upon the picture, but the story is interrupted at a critical point and in the third test the child is required to finish the story for himself. When this has been done a printed continuation of the story (Part B) is given him to read at his normal rate. This is followed by a test of verbal memory of what has been read. In the fifth test, the printed continuation of the story (Part C) is also given him to read, but this time he is cautioned that he must read so carefully that he can reproduce what he has read. The reproduction of this part of the story immediately follows. The concluding portion of the story (Part D) is then *read to* him and again the child is asked to reproduce the subject-matter. Test 1, which occurs on the last page, is called a test in handwriting, but needs considerable explanation. A portion of the test is shown in Fig. 1. It will be noted that the activity in the test is that of copying letters arranged in nonsense groups of five letters each. The actual letters used are those which occur in the dictation exercises, Test 2, and the frequency of occurrence of each letter is the same as in that test. As far as handwriting is concerned, the situation closely resembles that of the formal writing-lesson in which a copy is used. For this reason, this test has the *least* value of any for purposes of measuring handwriting. It does measure, however, very effectively, the rate of motor activity of the child, and his response to a strange situation involving no thought-content. It serves the further purpose of disposing of the question of handwriting. In this test, if anywhere, the child affects the mannerisms and style of the handwriting instruction, but, *having been tested in handwriting*, he writes normally and naturally in the tests which follow. In other words, comparison of the quality of handwriting in Test 1 with the average score made in Tests 3, 5, and 6 reveals any lack that there may be of transfer of the abilities generated by the handwriting lesson to the handwriting of a child when serving its normal function, that of recording thought.



"Measure the efficiency of the entire school, not the individual ability of the few."

R COURTIS STANDARD TESTS R R.

English. Test No. 1. Handwriting

SCORE
Number..................
Quality..................

Copy as many letters as possible in the time allowed. Write at the highest speed at which you can make well-formed, legible writing.

etaor hesin ldetc agoer

hysie tnlad eaurh wesib

faetn olreh dgsae timno

eruhl eatsd cieow ruhet

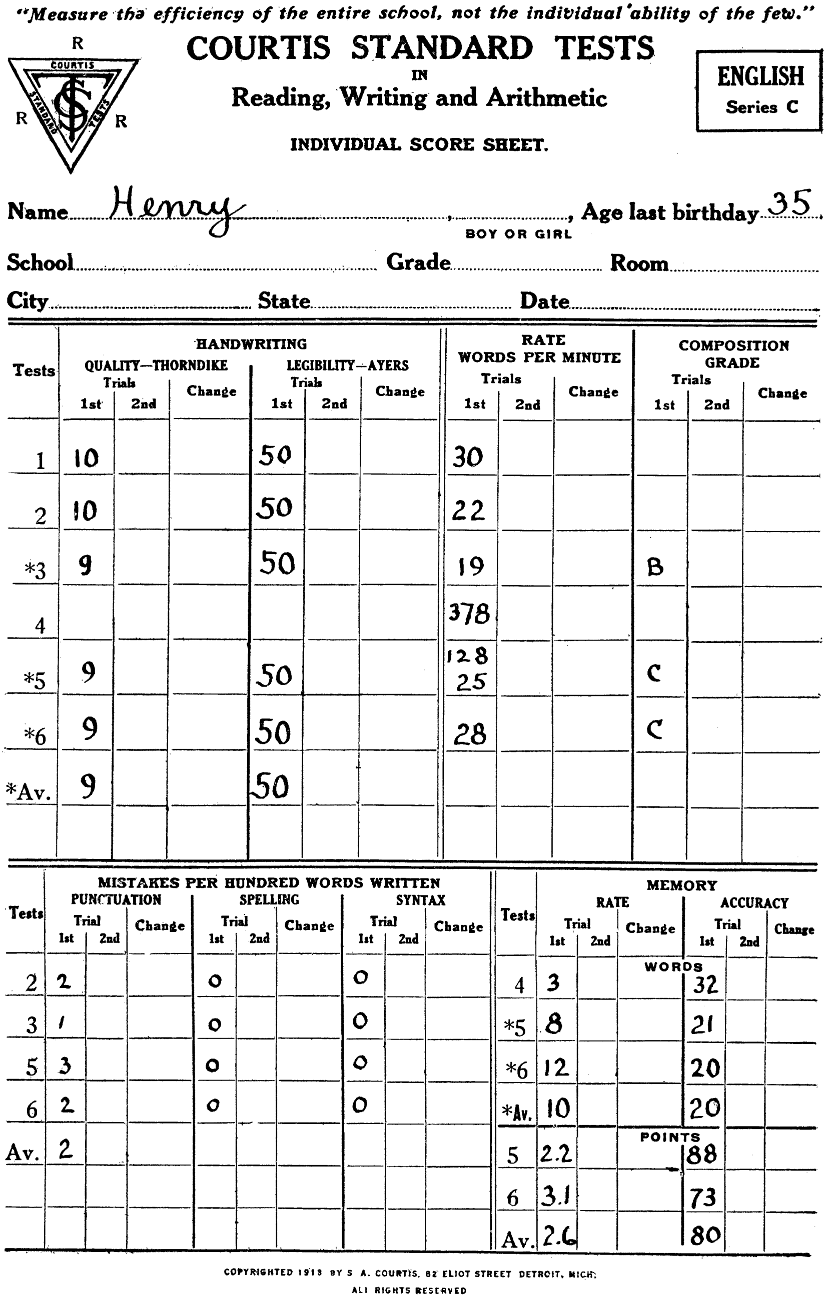
FIG. 1.—Portion of Test 1. This test is called a test of handwriting, but really measures rate of motor activity.

The tests are scored as to quality and legibility of handwriting by the Thorndike or Ayres scales, as to number of words read or written per minute, as to the quality of the original story and of the reproductions written, following the methods adopted by Rice, Bliss, and others, as to mistakes per hundred words written in punctuation, spelling, and syntax, as to the rate and accuracy of memory of both words and sense.

The details of the scoring and of the provisions taken to insure uniformity are too many to be given here. Two of these, however, need special comment. The question is often asked, why the Hillegas scale for measurement of English composition is not used. The answers are two in number. For one thing, the writer has not found the Hillegas scale a practical tool for classroom use. In the hands of a person of sufficient training, it affords a valuable objective basis of comparison of one type of composition with another. For instance, in comparing the products of future editions of the Courtis English tests with the present edition, the values of the standards in English composition would be equated by means of the Hillegas scale, but for measurements of school efficiency in teaching English composition it seems to the writer more practical to attempt to devise a scale, all the samples of which shall have a uniform subject-matter. This the products of the present series of tests will yield, and the construction of such a scale is one of the underlying reasons why the tests are issued in the present form. The second answer must be that the purpose of the tests is not so much to measure the abilities of the children in English composition as it is to determine the variation in judgment of teachers in various parts of the country, and the factors entering into such judgment. If as a result of giving these tests a thousand sample papers are secured all of which, in the judgment of the teachers scoring them, are "Excellent," and if these are compared with other groups of one thousand papers each, scored by the teachers "Good," "Fair," "Poor," "Failure," respectively, all on material uniform in content and obtained under uniform conditions, it will be possible to determine with great exactness what weight should be given to mistakes in spelling, writing, style, etc., in determining the grade of a paper. This would reduce the formulation of standards in English composition to a scientific objective basis. For this reason the teachers are asked to score the papers as to English composition according to *their own standards.*

A second point that needs discussion is the measure of spelling, punctuation, and syntax. The tests are in no sense spelling-tests, for instance, but they do measure spelling-efficiency. For if in one school under certain conditions 10 per cent of the eighth-grade children are able to write 100 words without making a single mistake, while in another school, under precisely the same conditions and on the same subject-matter, 53 per cent of the children in the eighth grade are able to write 100 words without a single error, it is evident that the efficiency of the instruction in spelling, other things being equal, is much greater in the second school than in the first. Tabulations of the returns from many schools will, therefore, yield standards of efficiency in spelling, punctuation, and syntax.

As a concrete illustration of the types of results secured from these tests, in Fig. 2 are given the scores made by an adult. This man, a supervisor in arithmetic in a school system in Michigan, was able to make a better score in the formal handwriting than he did in the later tests. The average quality in his handwriting by the Thorndike scale was judged to be that of sample 9, and his legibility equal to that of sample 50 on the Ayres scale. In the simple copying of groups of letters the individual wrote at the rate of 30 words per minute. In reproducing material which he had just read or which had been read to him, he wrote at the rate of 25 words per minute, but in the original story his rate fell to 19 words per minute. This fall in score is significant as revealing a difficulty which this person experienced, either in creating the ideas to write, or in putting into proper form the story he wished to express. His normal rate of reading was high—378 words per minute—but when reading with care for reproduction his rate was but 128 words per minute. The quality of his original story was "Good," as judged by the standards of the teacher who scored the paper, but his reproductions were only "Fair." He made on the average of two mistakes in punctuation per hundred words written, and none in spelling and syntax. His verbal memory is poor as shown by the very low rate of three words per minute in Test 4 and by the low accuracy of his results. This fact is again brought out in Tests 5 and 6. Here his rate is higher but his



"Measure the efficiency of the entire school, not the individual ability of the few."

COURTIS STANDARD TESTS
IN
Reading, Writing and Arithmetic

INDIVIDUAL SCORE SHEET.

ENGLISH
Series C

Name Henry , BOY OR GIRL , **Age last birthday** 35 .

School .... **Grade** .... **Room** ....

City .... **State** .... **Date** ....

Tests	HANDWRITING: QUALITY—THORNDIKE Trials 1st	2nd	Change	LEGIBILITY—AYERS Trials 1st	2nd	Change	RATE WORDS PER MINUTE Trials 1st	2nd	Change	COMPOSITION GRADE Trials 1st	2nd	Change
1	10			50			30					
2	10			50			22					
*3	9			50			19			B		
4							378					
*5	9			50			128 25			C		
*6	9			50			28			C		
*Av.	9			50								

Tests	MISTAKES PER HUNDRED WORDS WRITTEN: PUNCTUATION Trial 1st	2nd	Change	SPELLING Trial 1st	2nd	Change	SYNTAX Trial 1st	2nd	Change
2	2			0			0		
3	1			0			0		
5	3			0			0		
6	2			0			0		
Av.	2								

Tests	MEMORY: RATE Trial 1st	2nd	Change	ACCURACY Trial 1st	2nd	Change
			WORDS			
4	3			32		
*5	8			21		
*6	12			20		
*Av.	10			20		
			POINTS			
5	2.2			88		
6	3.1			73		
Av.	2.6			80		

COPYRIGHTED 1913 BY S. A. COURTIS, 82 ELIOT STREET DETROIT, MICH.
ALL RIGHTS RESERVED

FIG. 2.—A sample record showing scores made by an adult.

accuracy correspondingly lower. It is to be noted that he responds more readily to auditory stimulus than he does to visual. In spite of his poor verbal memory, however, his comprehension of what was read to him is high. He was able to reproduce 75 per cent of the essential subject-matter of the two passages, after a single reading.

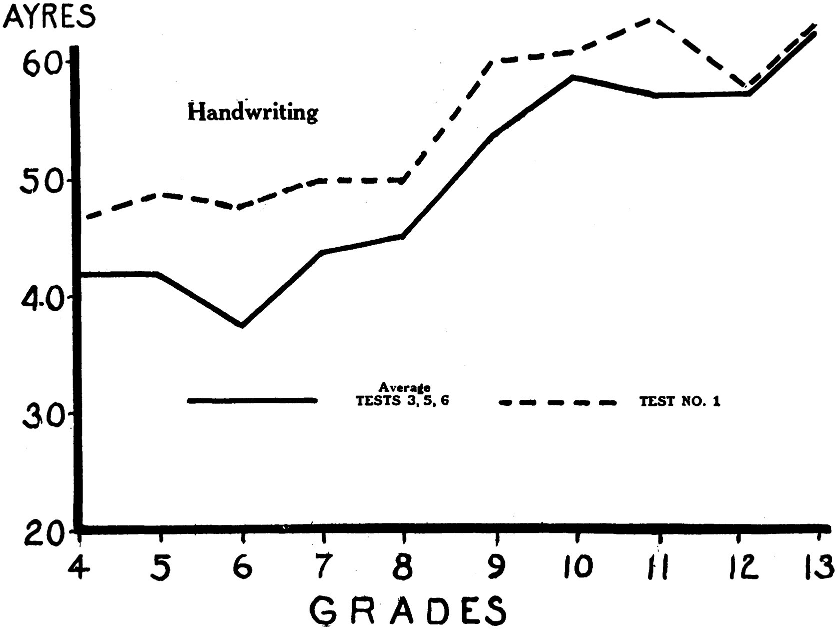


FIG. 3.—Median scores by grades in handwriting. Papers scored by the Ayres' scale. Solid line, average of median scores, Tests 3, 5, and 6. Dotted line, median score, Test 1, formal handwriting. The difference between the two curves was caused by the attempt to make good copy in the formal test.

It will be evident at once that the tests measure a number of very simple yet fundamental aspects of the ordinary products of school instruction in various phases of English work. The essential element in the interpretation of such scores, however, is standards based upon a sufficient number of cases for each grade to afford reasonable norms. These are, of course, entirely wanting at the present time, but in the graphs which follow, the results from the measurements of a single school are given. It should be particularly noted that the conditions of examination were kept uniform from the fourth grade through the high school (five-year course, hence grade 13) and that the scoring was done by one person in a uniform manner.

In Fig. 3 are given the median scores by grades in handwriting. It will be seen that the curve declines at the sixth grade, then rises to a maximum of 62 on the Ayres scale. This marks the efficiency of the handwriting instruction in this school as low, a fact which the school authorities had already recognized. The superiority of grades four and five over six indicates the effect of recent changes in the course of study with definite provision for handwriting instruction in place of the incidental training of previous years. The differences between the curves for Test 1 and the average of Tests 3, 5, and 6, differences which are fairly constant in grades 4, 5, 6, 7, and 8, indicate a lack of transfer of ability in these grades; in other words, the effort to produce good copy in a handwriting test leads to a better handwriting product than that which is obtained when the children are writing naturally. This should

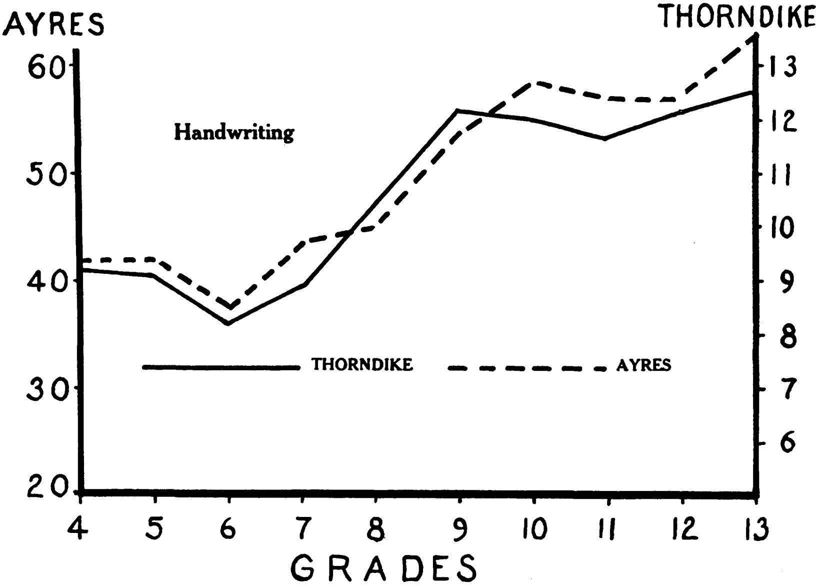


FIG. 4.—Comparison of the Development Curves for Handwriting as Determined by the Ayres and by the Thorndike scales.

serve to emphasize the need of *careful control of the conditions under which specimens are secured* in any measurement of efficiency of handwriting instruction.

In Fig. 4 are given for comparison the development curves for handwriting in this school, as measured by the Ayres and by the Thorndike scales. Except for grades 8 and 9, the differences in the two curves are slight. The fact that consistent *relative* results may be secured when samples of handwriting are measured by two different scales at different times is proof to the writer that objective measurement of handwriting rests upon no uncertain basis.

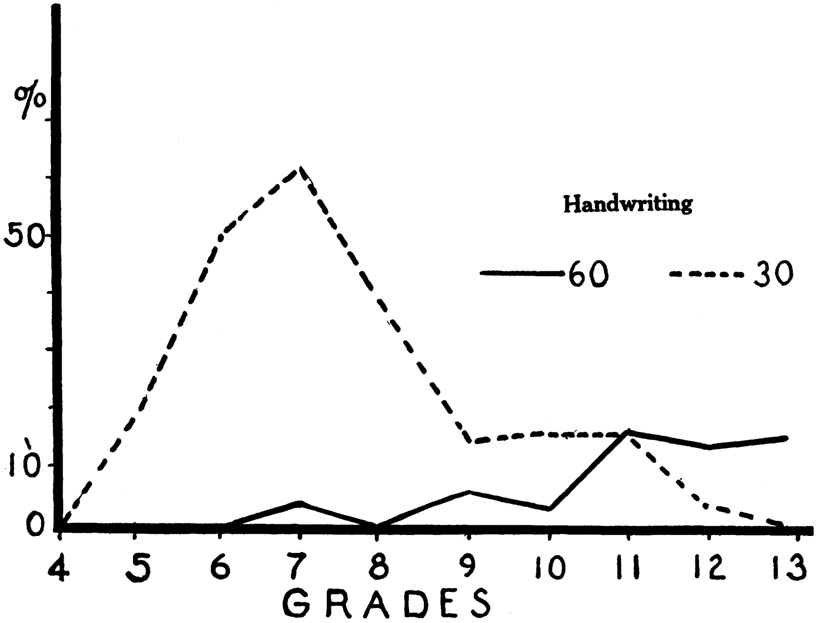


FIG. 5.—Efficiency curves for handwriting. Solid line shows the percentage of the membership of each grade that equals or exceeds sample 60 on the Ayres scale.

From the point of view of efficiency of instruction, however, median scores should be replaced by the percentage of the class membership that reach or exceed the desired standard. In Fig. 5 the heavy line shows the percentage of each grade's membership which equals or exceeds sample 60 on the Ayres scale. As noted before, the efficiency of the handwriting instruction in this school is very low, but at least the school now has an objective basis of comparison of present results with those which will have been reached when the present fourth-grade class shall have become Seniors. It will be remembered that a comparison of a school with itself or comparisons of methods from school to school was one of the fundamental purposes of such standard tests. In similar fashion the dotted line in Fig. 5 shows at once the percentage of each class membership whose writing equals or falls below sample 30 on the Ayres scale. Grades 5, 6, 7, and 8 are thus seen to have the greatest need for improvement.

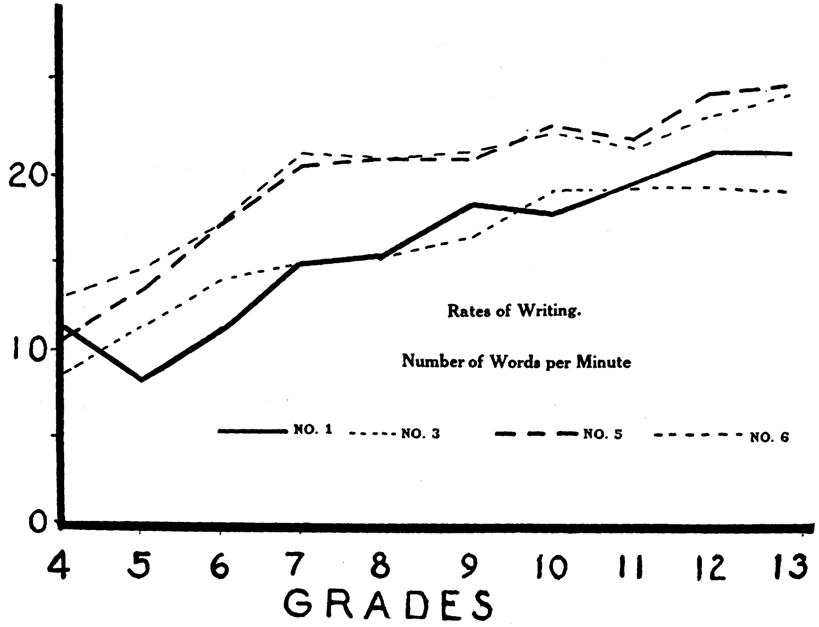


FIG. 6.—Development curves for rates of writing in Tests 1, 3, 5, 6.

In Fig. 6 are given the development curves for rate of writing, the number of words written per minute in Tests 1, 3, 5, and 6. It should be noted that curves for Tests 1 and 3 closely agree, as do those of Tests 5 and 6, but that the second set are considerably higher than the first. This means that slowing-down of speed of writing in Test 1 in the effort to produce good copy was about equal to the effect of the effort put forth in composing the original story in Test 3. In other words, the difference between the two sets of curves measures the retarding influence of mental effort. Eighth-grade children able to write at the rate of 22 words per minute when reproducing a story that they have read, or that has been read to them, were able to write but fifteen words per minute when they were forced to supply their own ideas, or when they are taking particular care to make legible writing. In this ability as in so many others, the period of rapid growth ends at the seventh grade.

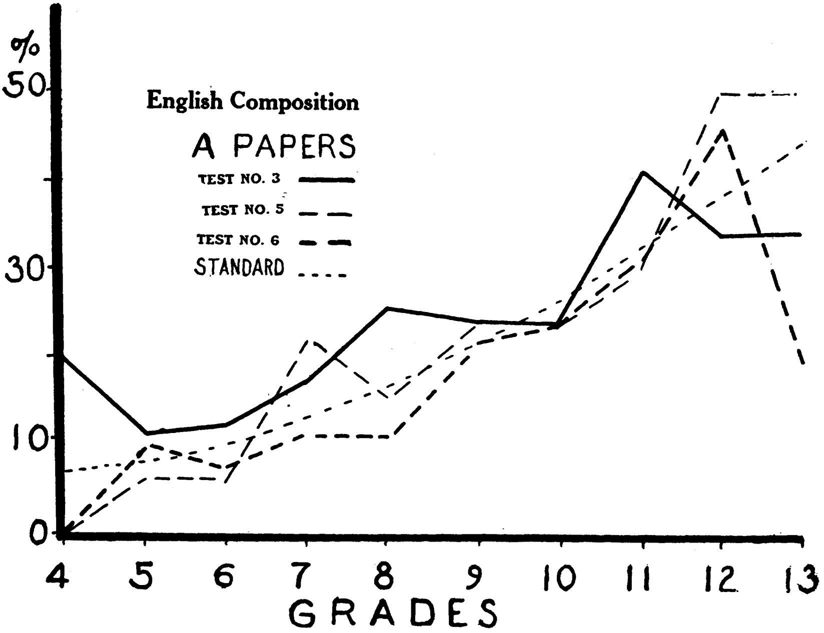


FIG. 7.—Development curves for the ability to write an "A" (Excellent) paper in the various tests.

In Fig. 7 are drawn the curves showing the percentage of A ("Excellent") papers in each grade for various tests. It is interesting to note that in nearly every grade the curve for the original story is higher than the curves for either of the two reproductions, as it should be by the doctrine of interest. The dotted line marked "standard" represents, for this school, the efficiency of present instruction. That is, the efficiency of training in English composition at the present time is such that at the eighth grade but 13 per cent of the children are able to write A, or "Excellent," papers.

Fig. 8 shows the curves based upon the mistakes in punctuation, spelling, and syntax per hundred words written. Development of these abilities, like many others, is most rapid in the early grades but continuous through to the end of the high-school years.

More interesting, however, are the efficiency curves given in Fig. 9. These are based upon the percentage of papers in each grade which had between one and four mistakes in punctuation per hundred words, less than one mistake in spelling and in syntax per hundred words. Note the extremely rapid rise of the curve for punctuation and the more gradual development of the curves for spelling and syntax. It is interesting to note that while at the eighth grade but *one child in ten* is able to write 100 words without making a mistake in spelling, by the thirteenth grade the number has risen *to 55 per cent.* In this school the greater development of these abilities comes, not during the years when provision

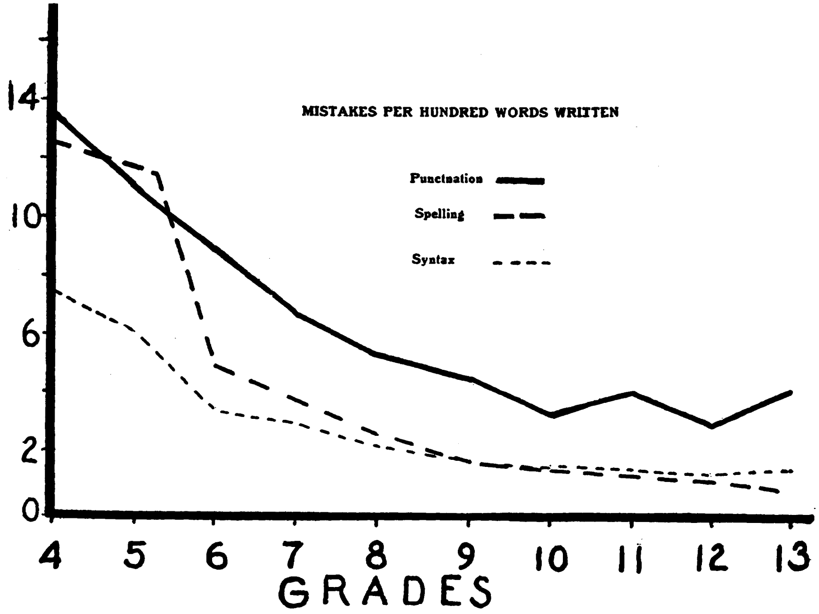


FIG. 8.—Development curves for punctuation, spelling, and syntax as determined by the number of mistakes in each per hundred words written.

is made for active study of spelling (grammar grades), but during high-school years *when the corrective work is incidental.* If these results are confirmed by the tests of other schools, it probably means that here, as in arithmetic, training on the specific abilities of spelling, punctuation, and grammar, is nearly valueless and does not transfer, that our first concern must be to give the children experience in thinking, expressing, and recording their thoughts and later perfect the mechanical elements through repeated use.

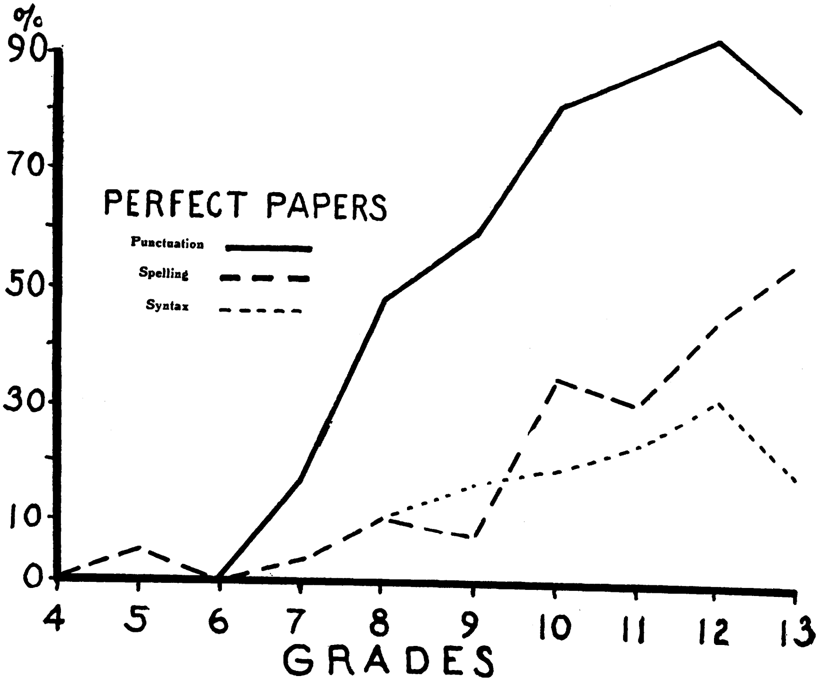


FIG. 9.—Efficiency curves for punctuation, spelling, and syntax. The curves indicate the percentage of each class membership which made but from 1 to 4 mistakes in punctuation, and less than 1 mistake in spelling and syntax, per hundred words written.

The curves in Fig. 10 give the median scores for rates of reading. It will be seen that the curve for normal reading rises to a high value and does not reach its maximum until the high-school years. The curve for careful reading, on the other hand, is practically constant from the sixth grade on. This probably means that the rate and character of one's serious reading is fixed in the early school life. The result, however, needs to be confirmed by the tests of many school systems; it may be due to purely local conditions. If this type of curve should prove general, it should be possible eventually to define the ability to read in objective terms, i. e., a child would have standard ability in reading as soon as it could read 230 words per minute of a simple narrative printed in the size of type, length of line, openness of spacing represented by Test 5. To this must be added the qualifying clause, "and be able to reproduce about 50 per cent of the words and 75 per cent of the ideas after a single reading."

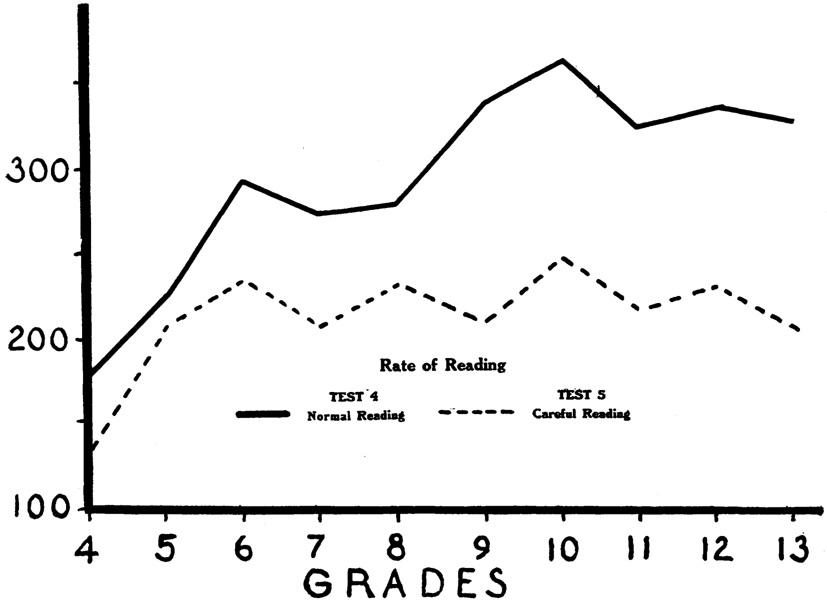


FIG. 10.—Development curves for normal and for careful reading, based upon the number of words read per minute.

The results of the tests prove that the range of individual ability in reading, writing, etc., is as great as that revealed by the arithmetic tests. The performance of the more able child in any one class is usually from two to three times that of the less able child.

In the upper half of Fig. 11, for instance, the distribution of the 175 children in grades nine to thirteen are shown for Test 4 (solid line) and Test 5 (dotted line). The scale along the base of the figure represents the number of words read per minute, while the scale at the left shows the frequency of the different scores. In each test a group of 50 children exceed the limit of the test. (The experimental work upon which the lengths of the various tests were based was done mainly with adults.) Had the tests been long enough, the ability of these pupils would undoubtedly have varied from 200 words per minute to 500 words per minute for normal reading, and from 60 words per minute to 400 words per minute for careful reading. When the vital nature of ability to read and understand is considered, this range of ability to read the simplest prose becomes very significant. Precisely similar results

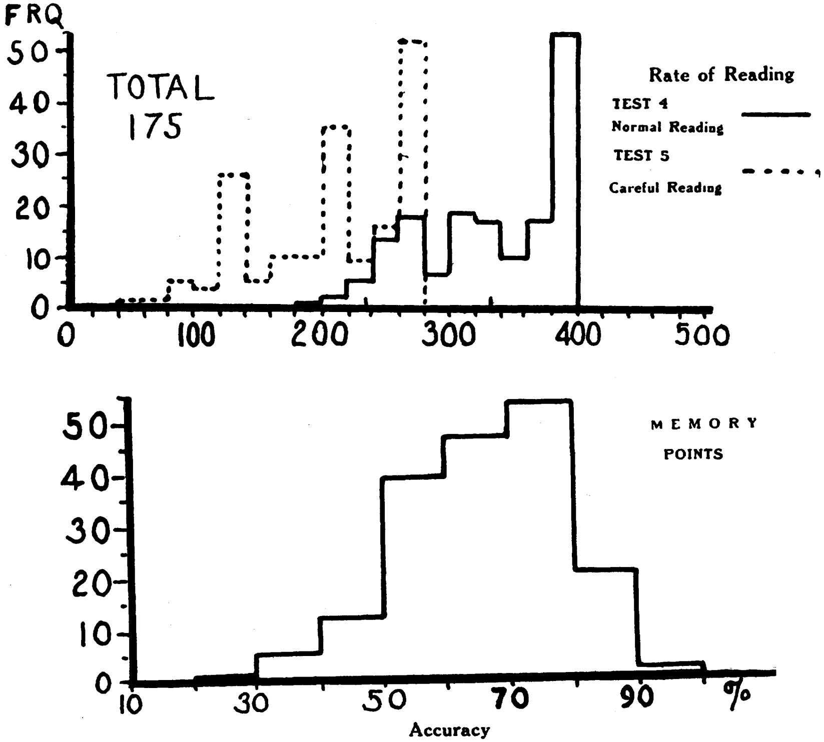


FIG. 11.—Distributions of individual scores of 175 high-school pupils as samples of the great range of individual variation in all tests. Abilities selected, the number of words read per minute in normal reading and in careful reading, and the number of ideas remembered of what was read. The more able children have scores from two to three times as large as those of the less able.

are shown for memory in the lower half of the figure where the scale along the base line shows the accuracy with which the ideas read were remembered. It is evident from this one test alone that the character of the school product in reading is precisely similar to that in arithmetic, and the writer expects to find that the same general causes operate to prevent success, that the same factors determine efficiency, and that the same changes in methods of teaching will prove effective.

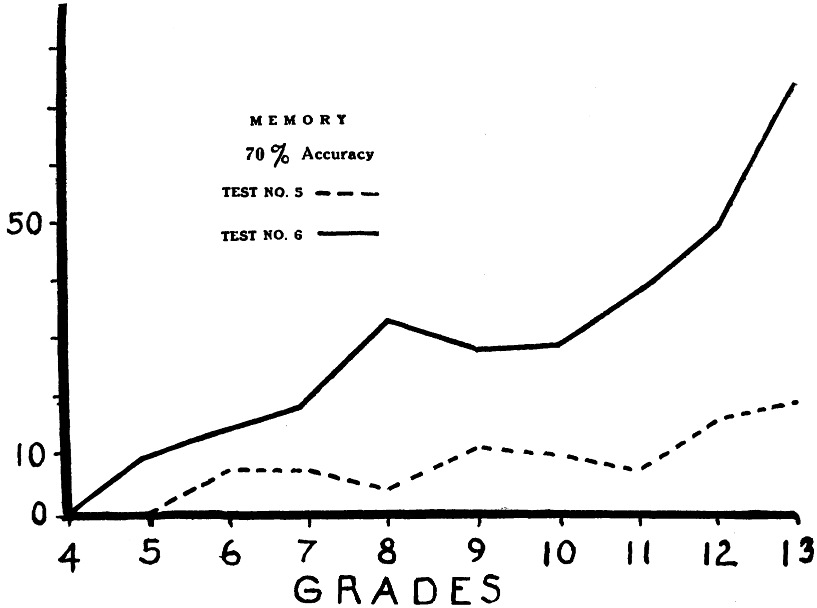


FIG. 12.—Curves showing the percentage of each class membership remembering 70 per cent or more of what was read (Test 5, visual stimulation) and of what was heard (Test 6, auditory stimulation). Note the rapid rise of the second curve, particularly in high-school years, and the very slight increase in the first curve.

The results shown by the curve in Fig. 12 were unexpected. The solid line is based upon the percentage of each class who remember 70 per cent or more of what was read to them (auditory stimulation), while the dotted line is a similar curve for the material the children read themselves (visual stimulation). It is evident from the figure that the children are steadily growing in their ability to remember the spoken word, while their ability to retain what they themselves read is much less and shows little increase during the period of school life. In this particular school, at least, the growth in the power of the visual stimulus is much less than has been supposed.

In Fig. 13 is given a graphic representation of the use of the products of such standard tests in dealing with individual children. On the basis of these tests and the requirements of the school it is possible to say that an eighth-grade child of standard ability should

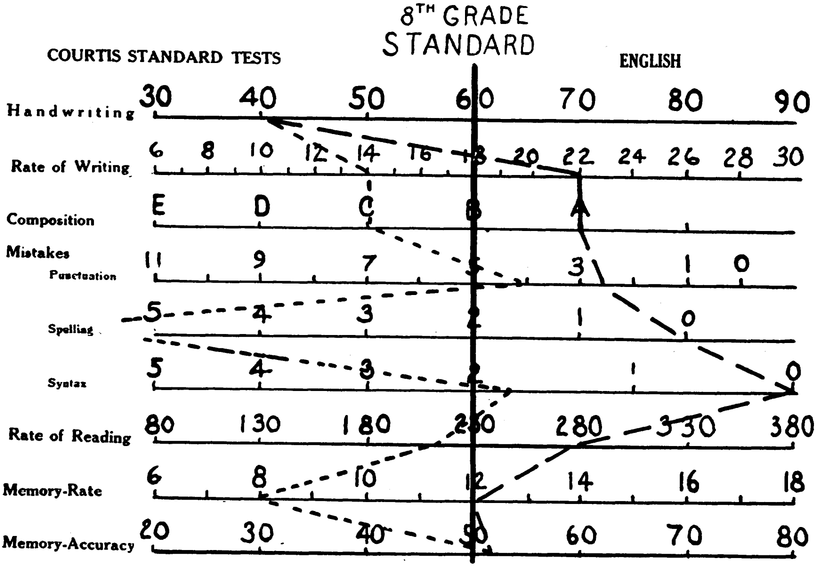


FIG. 13.—Comparative graph of standard and individual scores in the various tests. A suggestion of the possible use of scores in such tests for diagnosis of individual defects.

be able to write an original story at the rate of 18 words per minute and that legibility of the writing should be 60 on the Ayres scale, and the quality of the story "Good"; that in the story there should not be more than five mistakes in punctuation per hundred words, two in spelling, and two in syntax; that in careful reading of the given material under standard conditions, the rate should be 230 words per minute, and that in the reproduction of the material read, 12 of the original words should be used per minute and that these words should constitute 50 per cent of the words used in the reproduction. The graph shows the actual scores made by two members of the same class. One individual—represented by the heavy broken line—wrote 22 words per minute, but the quality of her handwriting was only 40. Her story, however, was excellent. She made but 2.7 mistakes per hundred words in punctuation, none in spelling and syntax. She was able to read the test material at a high rate of speed and could reproduce what she had read at the standard rate and with better than standard accuracy. The other individual made the same scores in handwriting, although her rate was but 14 words per minute in the original story and the quality of the story itself but "Fair." Her control of punctuation and syntax was above standard but she was a poor speller. Her rate of reading was nearly normal, but she was slow in reproducing what she had read, although what she actually did write was up to the standard in accuracy. It is evident from these curves that both individuals need additional training in handwriting; that the second one needs also special assistance in spelling and special practice in the control of the mechanism by which the ideas are expressed.

The work in arithmetic has proved that it is not possible radically to change the efficiency of present methods *until the actual work assigned to each pupil is based upon his measured needs.* It should be evident from the foregoing discussion that the products from these tests in English will furnish objective standards that will serve both as goals for the guidance of teachers and pupils, and as a means of detecting the peculiar weaknesses of individuals. The co-operation of all teachers and superintendents interested in the experimental evaluation of such aids to classroom work will be appreciated.